

MURDER DATA ANALYSIS

A PROJECT CASE STUDY

Dena Coduri



THE PROJECT

Background

This is a personal, open exploratory project investigating homicides in the United States. The main data set is from the Murder Accountability Project (MAP), a nonprofit whose purpose is to educate the public about the unsolved murder problem in the U.S., and to help police departments solve cold cases by providing data from the FBI/local governments.

Objective

My purpose with this project is to source, clean, and analyze the homicide data set as an exploration, constructing and answering interesting questions along the way. I also investigate if the homicide problem is as big as the MAP claims it to be. Important insights are included in a final deliverable as a Tableau story board.



THE ANALYSIS

Steps



The data used for this analysis is publicly available from the [Murder Accountability Project](#), supplemented with population data from the [U.S. Census Bureau](#)



Software used –
Python, Tableau,
Excel, Word



Data cleaning performed including removal of duplicate rows, imputing median ages into missing cells, and fixing discrepancies.

- Source and clean/prep the data, documenting each step in Word
- Derive new columns and investigate subsets of the data
- Create a list of questions to explore
- Answer the research questions using Python visualizations and statistical methods
- Merge the murder data with the cleaned Census population data from a previous project
- Create a story board in Tableau with the most interesting insights, answering any remaining research questions using Tableau

THE QUESTIONS

I compiled the following list of questions to explore for this project:

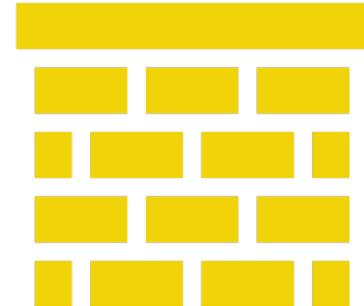


- How has the number of homicides fluctuated over the years?
- Are homicides more common in certain states?
- Are wives more likely to murder their husbands, or are husbands more likely to murder their wives?
- Is there a correlation between the number of victims and the age of the offender?
- Which gender is most likely to be a victim in a Lover's Triangle circumstance?

THE OBSTACLE

One of the data sets from the Murder Accountability Project included data from cases where there were multiple victims and/or offenders. It was problematic because the same case was listed multiple times in these instances: one extra entry for each additional person involved. That made it difficult to analyze because the same situations were listed multiple times and would have been double counted.

I chose to narrow my data frame and simply analyze the homicides where it was known that there was a single victim and a single offender, which made up over 55% of the cases.

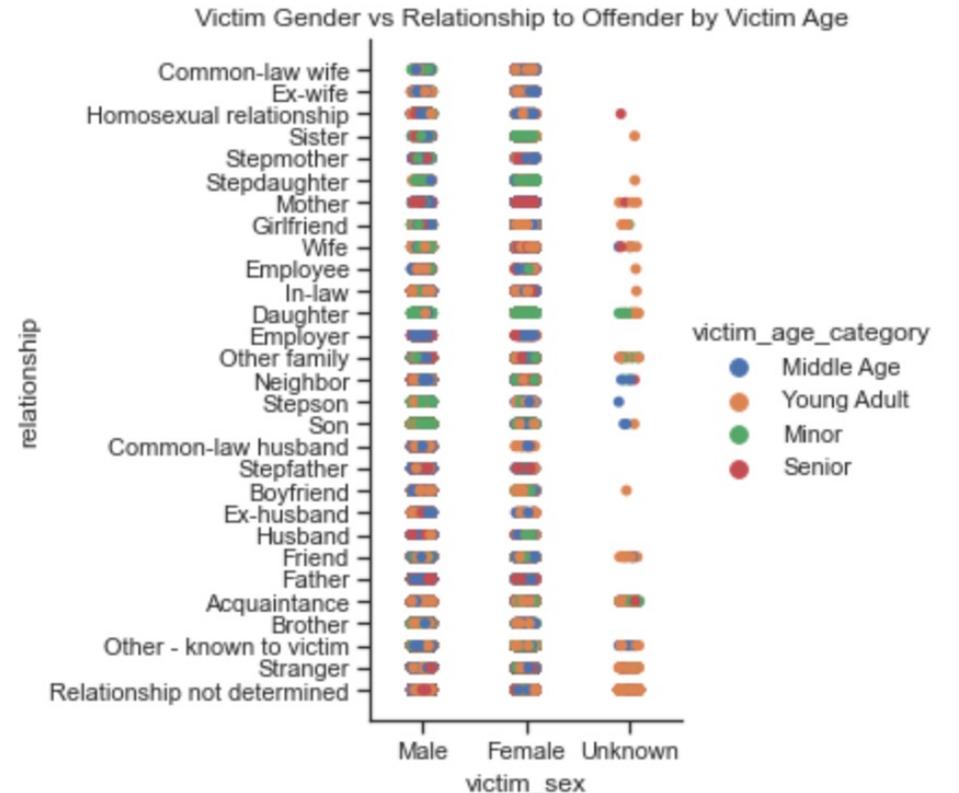


THE EXPLORATION

I created a hypotheses based upon this categorical plot I made in Python using Seaborn:

- Offenders who murder their daughters and sons are more likely to do so when the victims are minors.

Using a proportional area chart in Tableau, I was able to see that this hypothesis was indeed true. Far more people who die at the hands of their parents are under the age of 18.



THE DELIVERABLE



An Analysis of Homicide in the United States

How much is known about U.S. homicides? | What are the trends in homicide numbers? | What circumstances surround homicides? | How does age factor into homicides? | The frightening truth about the unsolved murder problem | Analysis limitations and next steps

How much is known about the circumstances surrounding homicide cases?
Is it true that one-third of all murders in the United States go unsolved?

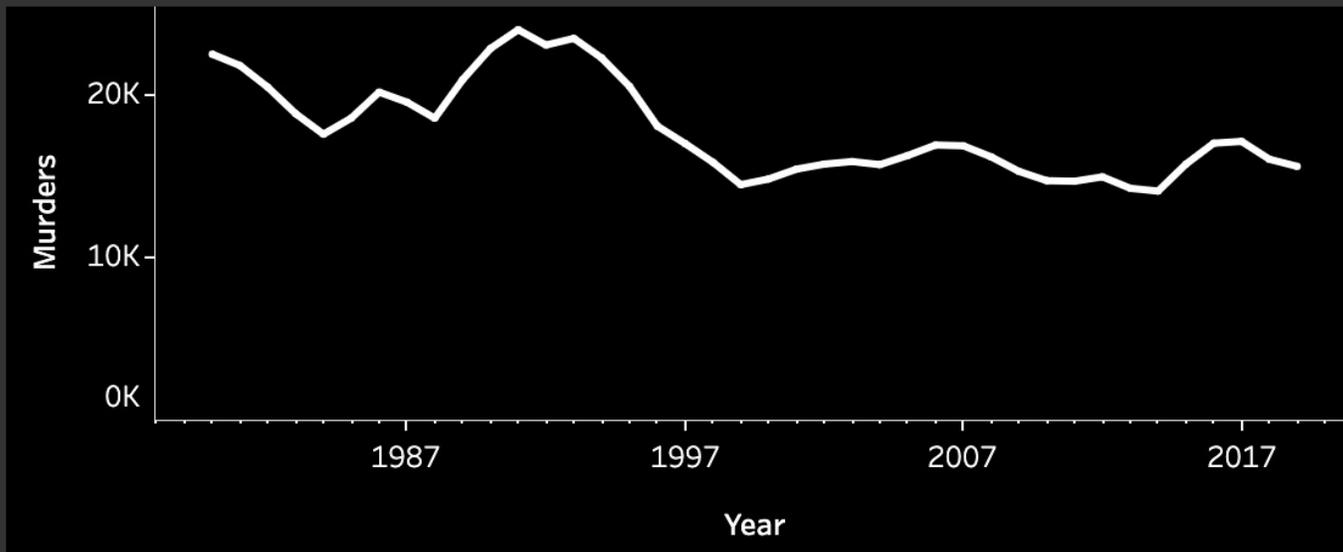
This exploratory project looks at homicide trends in the U.S. and uncovers some of the grisly truths about the circumstances and situations surrounding these murder cases. The homicide data used for this project is publicly available from the Murder Accountability Project, and the incorporated population data comes from the U.S. Census Bureau.

About the homicide data:
The Murder Accountability Project (murderdata.org) is a nonprofit group that collects and publishes homicide information from the FBI, supplemented with information from local governments. The mission of the organization is to help local police departments solve cold cases and to educate the public about the country's unsolved murder problem.

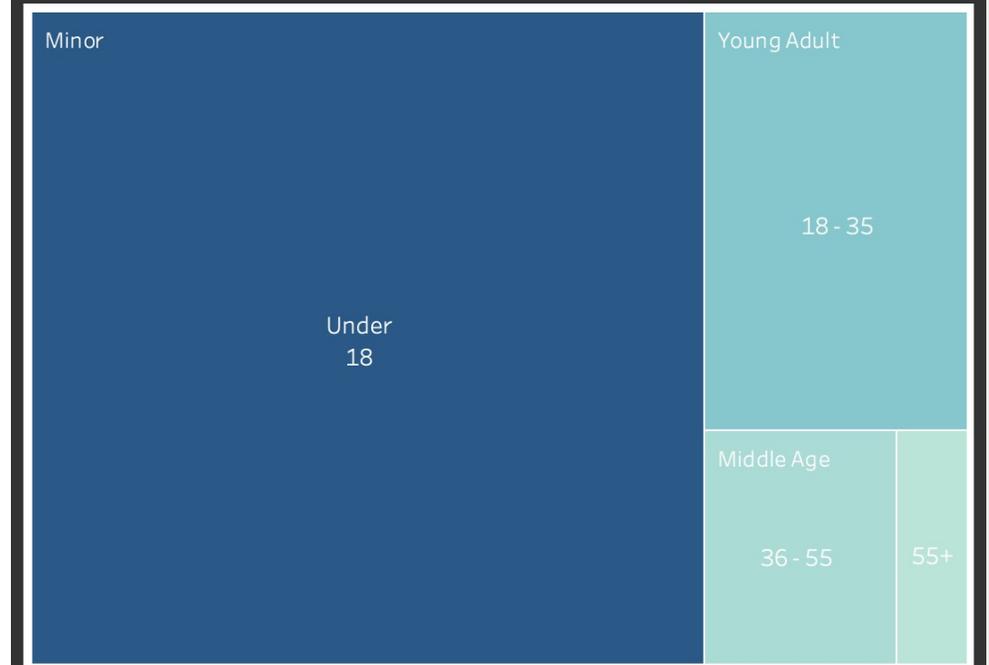
The deliverable is a Tableau story board, with different dashboards for insights regarding trends, circumstances, age, the unsolved murder problem, and next steps.

THE VISUALS

Total Number of Homicides
1980 - 2019



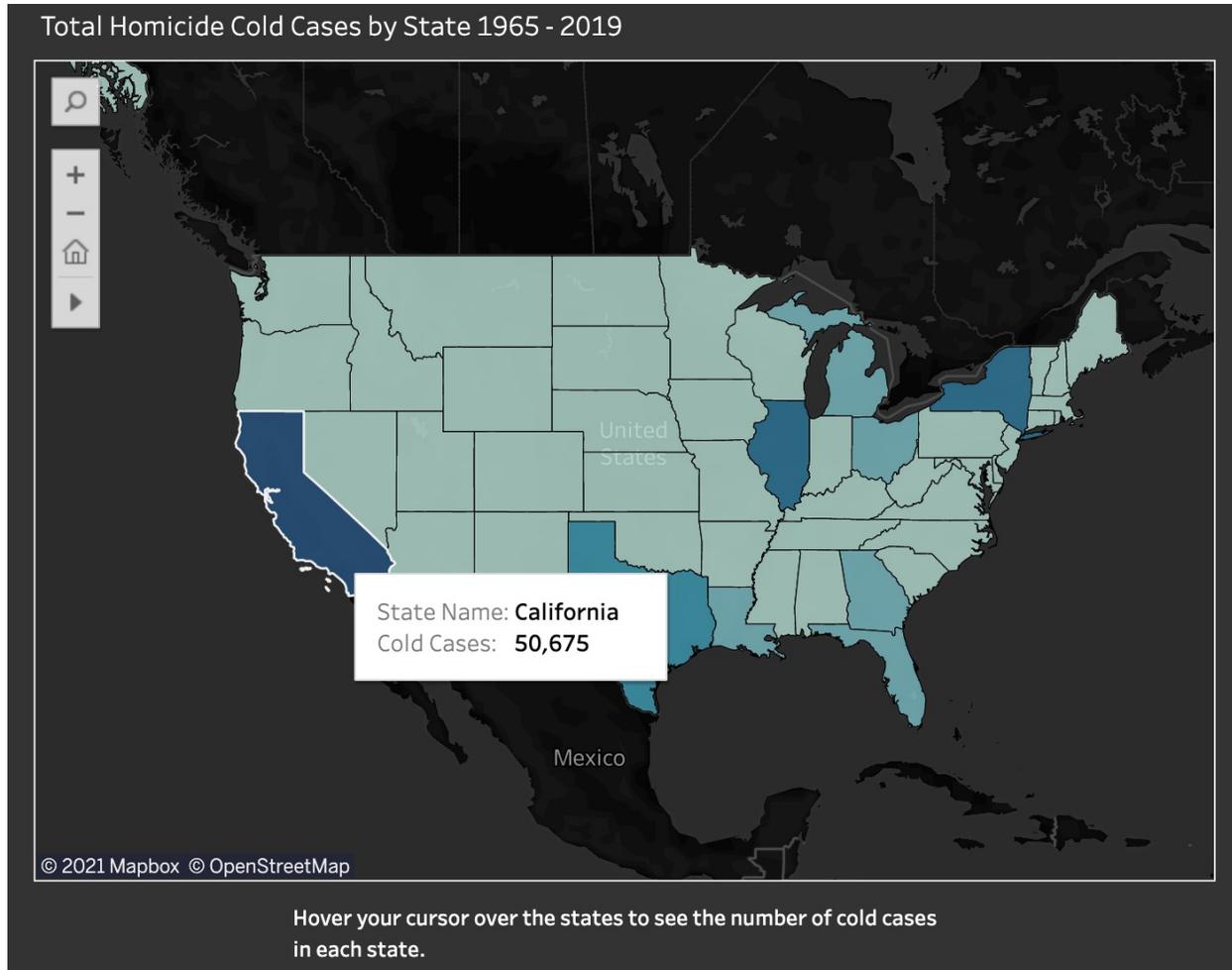
Ages of Victims Killed by a Parent 2015 - 2019
(Single Victim/Single Offender Situations)



Many visuals were created for this project, including heatmaps, scatterplots, line graphs, and proportional area charts. More visuals were created in Python during the exploratory analysis phase but were not used in the final deliverable.

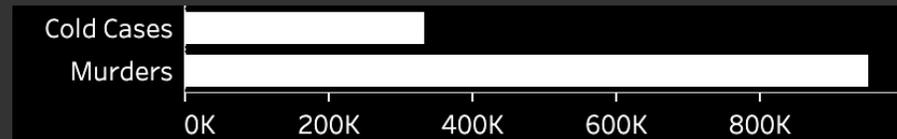


THE INSIGHTS



35% of all homicide cases between 1965 and 2019 remain unsolved.

Total Number of Murders and Cold Cases in the U.S. 1965 - 2019



The unsolved murder problem in the United States is real. One-third of all homicide cases remain unsolved. There are over 50,000 unsolved homicides in California alone.



[Link to Tableau Story Board](#)
[Link to GitHub Repository](#)

THE RETROSPECTIVE

- There were a lot of questions that I did not get to explore. To add upon this project, I would like to look at research questions regarding homicide trends grouped by political parties of the corresponding state, and homicide trends grouped by race.
- I would also like to analyze the homicide cases where there was more than one offender and/or victim. These were challenging to analyze because of the way they were recorded, but with more time on my hands, I'm sure I can reorganize the data and find some insights there.
- This project was the first in which I had to source my own data. I enjoyed the process of scouring the web for open source, reliable, and interesting data to analyze. Going forward, I would like to try my hand at data mining.

